



APR 23 1999

WASTE FILE

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES
CHAPTER #LL- 1

MEMORANDUM FOR Barbara M. LoPresti
Chief, Technologies Management Office

From: Howard Hogan *Howard Hogan*
Chief, Decennial Statistical Studies Division

Prepared by: Steven P. Hefter *SPH*
Long Form Sample Design and Estimation Team

Subject: Sample Tolerance Check Specifications for Census 2000

I. Introduction

This memorandum provides the specifications for the Sample Tolerance Check (STC), to be conducted in List/Enumerate (L/E) areas for Census 2000.

This procedure is used to determine if there is a statistically significant difference between a simplified long form estimate of total population within two consecutive L/E Assignment Areas (AAs) that were completed by a single enumerator, and the total population count from the initial phase, for the same two AAs. This procedure is intended to detect bias in the long form sample stemming from enumerator error. Enumerator bias in L/E AAs arises when enumerators ignore the systematic sampling procedure and administer the long form questionnaire to a disproportionate number of vacant housing units (HUs), or to HUs containing predominately small or predominately large families, thus adding a bias to the long form sample.

The specifications for determining the Census 2000 long form sample will be provided in a future memorandum.

These specifications should be used to flowchart the process, to generate further discussion on requirements, to identify and finalize the record layouts of input and output files, and to write computer software to implement the methodology. During and after a testing phase, it is likely that changes to the specifications will be necessary to reflect final data processing plans and to account for unanticipated issues.

Any comments or questions regarding these specifications can be directed to Philip Gbur

at (301) 457-4206 or Steven Hefter at (301) 457-4082.

II. Overview

For L/E areas, address listing books will be prepared with the appropriate form type assigned to each line. Enumerators will conduct the L/E operation and return the books to their Local Census Office (LCO). Batches, consisting of two consecutive AAs completed by a single enumerator, will be created for the STC. The LCO staff will capture some basic data from the listing books and run the STC computer software developed by the Technologies Management Office (TMO). If a batch fails the STC, both AAs within the batch will be resampled independently and returned to the field for additional interviewing during the Coverage Improvement Follow-Up (CIFU) operation. In some instances, AAs that pass the STC, but are sampled at a lower rate than designated will be subject to a supplemental sampling of addresses.

See Attachment 1 for a flowchart of the process.

The following section provides the detailed specifications for batch creation, testing and if necessary, the resampling of the AAs, which are to be incorporated into the STC software.

III. Methodology

The STC employs a standard, two tailed statistical hypothesis test with the significance level (α) = 0.01. The hypotheses for the STC are:

H_0 : The weighted long form estimate of the total population in an L/E batch is *equal* to the 100 percent census population count in the L/E batch.

H_a : The weighted long form estimate of the total population in an L/E batch is *not equal* to the 100 percent census population count in the L/E batch.

- A. Create a batch from two consecutive AAs completed by an enumerator. Number each batch with an integer beginning with the number one. If an enumerator has completed only one AA or an odd number of AAs, treat the single AA as a batch.
- B. Perform the STC on each batch, after each AA within the batch has been completed. An AA is considered complete after all necessary field reinterviewing has occurred. For each batch, create a file with one record per census questionnaire. Create a variable for form type and a variable for the number of persons enumerated at each HU. Sort the file by: form type, block, and census identification number (ID). This file is the input to the STC.

C. Compute:

n_i = the number of HUs enumerated on a long form questionnaire within the i^{th} AA in the batch

N_i = the total number of HUs enumerated within the i^{th} AA in the batch

Note: Carry all non-integer variables to the sixth decimal place.

D. Compute:

1. The total number of HUs enumerated within the batch (N),

$$N = \sum_{i=1}^2 N_i \quad (i=1,2) \quad ; \text{ and}$$

2. The number of HUs enumerated on a long form within the batch (n),

$$n = \sum_{i=1}^2 n_i \quad (i=1,2)$$

Note: If $N < 10$ or $n < 7$ do not perform the STC for that batch.

E. Compute the observed sampling rate for the i^{th} AA as:

$$OSR_i = (n_i / N_i) \quad (i=1,2)$$

F. Define SR_i to be the sampling rate that was used in the i^{th} AA within in the batch. Determine SR_i ($i=1,2$) as follows:

1. If $OSR_i \geq 0.25$, then $SR_i = 1/2$; else

2. If $0.0 \leq OSR_i < 0.25$, then $SR_i = 1/6$

G. Determine the following:

$Y_{i,j}$ = The number of persons enumerated at the j^{th} HU within the i^{th} AA in the batch,

Y_i = The total number of persons enumerated within the i^{th} AA in the batch,

$$Y_i = \sum_{j=1}^{N_i} Y_{i,j}$$

$Y_{L,i}$ = The total number of persons enumerated on long forms within the i^{th} AA in the batch,

SF_i = The number of short form HUs within the i^{th} AA in the batch,
 $SF_i = (N_i - n_i).$

\hat{Y} = The long form estimate of total population within the batch,

$$\hat{Y} = \sum_{i=1}^2 (Y_{L,i} / SR_i)$$

Y = The total number of persons enumerated within the batch,

$$Y = \sum_{i=1}^2 Y_i \quad (i=1,2)$$

$AASR_i$ = The sampling rate for the i^{th} AA within the batch.

H. Compute S^2 for the i^{th} AA as follows:

$$S_i^2 = \sum_{j=1}^{N_i} (Y_{ij} - (Y_i / N_i))^2 / (N_i - 1) \quad (i=1,2)$$

Note: If $S^2 = 0$ for *both* AAs within a batch, do not perform the STC. No resampling is required within the batch.

I. Calculate an estimate of the variance (V) of the long form estimate of total population within the batch as follows:

$$V (\hat{Y}) = \sum_{i=1}^2 ((1 / SR_i) - 1) \times N_i \times S_i^2$$

J. The Test Statistic (t)

The test statistic (t) used in the STC is the absolute difference between the long form population estimate and the 100 percent census count of the population divided by an estimate of the standard error of the long form population estimate.

$$t = | \hat{Y} - Y | \div \sqrt{V(\hat{Y})}$$

where:

$| a - b |$ = the absolute value of $a - b$.

K. The Test

1. Calculate the test statistic.
 - a. Compute \hat{Y} , $V(\hat{Y})$
 - b. Compute t .
2. Determine $t_{critical}$ based on the table in Attachment 2.
 - a. If $t \leq t_{critical}$, the batch has passed the STC.
 Check each AA within the batch.
 - i. If either $SR_i = AASR$, or $SR_i > AASR$, stop. No resampling is necessary for this AA.
 - ii. If $SR_i < AASR$, and $N_i > 9$, go to IV.A.
 - b. If $t > t_{critical}$, the batch has failed the STC. Go to IV.B.

IV. Resampling Procedures

Note: All resampling procedures occur within the individual AAs.

- A. Procedure for computing the TE for AAs passing the STC using the SR_i , with $SR_i < AASR_i$.

Due to the fact that the AA passed the STC using the SR_i , which is lower than the $AASR_i$, the initial sampling procedure was unbiased, although the sample size is smaller than designated. Choose a sample of short forms within these AAs to be redesignated as long forms, as follows.

1. Compute the TE, as:

$$TE_i = 1 / [(AASR_i - SR_i) / (1 - SR_i)]$$

2. Implement the sampling procedures in IV.C *only*, using the TE_i computed in IV.A.1. *The existing sample of long forms is retained.*

- B. Determine the take-every (TE_i) as follows:

$$TE_i = 1 / AASR_i$$

- C. Choose a sample of short forms to be redesignated as long forms as follows:

1. Generate a sequence of integers $\{L_x\}$ as follows:
 - a. Generate a random number, r , such that $0 < r \leq TE_i$;
 - b. Let $L_1 = r$; and
 - c. Starting with $x = 2$, calculate $L_x = L_{x-1} + TE_i$, while $L_x \leq SF_i - 1$
2. Compute $CEILING [L_x]$ ¹

Example: If $r = 1.2$, and $TE_i = 2$, then $L_1 = 1.2$. The generated L_x s would be the sequence: 1.2, 3.2, 5.2, 7.2, ..., $(L_{x-1} + 2)$. Therefore, the 2nd, 4th, 6th, 8th, ..., and the $CEILING [L_x]$ th short forms are chosen.
3. Assign the short form HUs, chosen in IV.C.2 above, for follow-up interviews with long form questionnaires.
4. Flag these HUs. This flag is *FLAG1*.

- D. Choose a sample of long forms to retain as follows:

Note: The long forms referred to here do not include those converted from short forms in IV.C.

1. Generate a sequence of integers $\{L_z\}$ as follows:
 - a. Generate a random number, q , such that $0 < q \leq TE_i$;

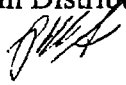
¹ $CEILING [X]$ is the nearest integer greater than or equal to X . For example, $CEILING [2.87] = 3$

- b. Let $L_1 = q$; and
- c. Starting with $z = 2$, calculate $L_z = L_{z-1} + TE_z$, while $L_z \leq n - 1$.
2. Compute *CEILING* [L_z]
3. Retain these long forms. Convert all other long forms to short forms.
4. Flag these redesignated long form HUs. This flag is *FLAG2*. The resampling is complete for the AA.

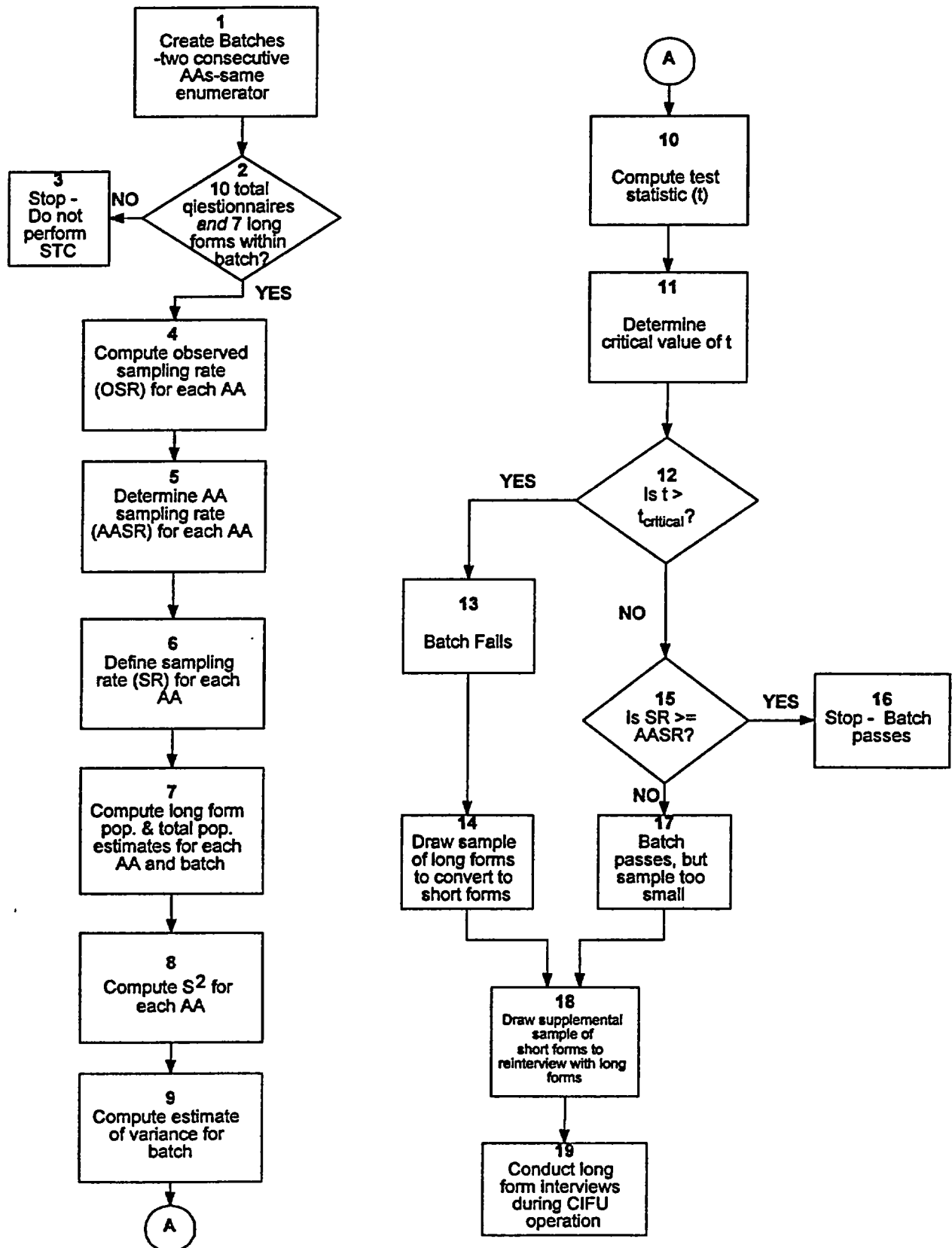
V. Test File

The Decennial Statistical Studies Division (DSSD) will provide a series of test files to the TMO for testing the implementation of the methodology provided in these specifications. The files will be provided in a format determined through discussions with the TMO. Provide the results of the test to the DSSD on paper or in ASCII format. The requested variables are given in Attachment 3.

cc: DSSD Census 2000 Procedures and Operations Memorandum Series Distribution List
 Long Form Sample Design and Estimation Team Distribution List

J. Dawson	(TMO)	
W. Ballew		
V. Gore		
P. Montgomery		
S. Walker	(DMD)	
K. Lind		
A. Quinlan		
C. Kahn		
K. Giesbrecht		
F. Borsa	(FLD)	
L. Minneman		
C. Kriesberg		
D. Sefton		
A. Pfeiffer	(GEO)	
G. Brooke		
R. Damario		
C. LeFevre		
P. McGuire	(DSCMO)	
R. Chilson	(DSSD)	
M. Tenebaum		
R. Feldpausch		
M. Hudson		

Sample Tolerance Check Flowchart



Values of $t_{critical}$ for Comparison to Calculated t Values

$n-1$	$t_{critical}$
6	3.707
7	3.499
8	3.355
9	3.250
10	3.169
11	3.106
12	3.055
13	3.012
14	2.977
15	2.947
16	2.921
17	2.898
18	2.878
19	2.861
20	2.845
21	2.831
22	2.819
23	2.807
24	2.797
25	2.787
26	2.779
27	2.771
28	2.763
29	2.756
≥30	2.576

Output for Verification

The following output from the Sample Tolerance Check test files is requested by the Decennial Statistical Studies Division (DSSD). Please provide the DSSD the following output upon the completion of the testing process, for each batch tested.

STC Test Output

1. The number of HUs enumerated on long forms within each AA (n_i)
2. The total number of HUs enumerated within each AA (N_i)
3. The observed sampling rate within each AA (OSR_i)
4. Sampling rate for each AA (SR_i)
5. Total number of persons enumerated within each AA (Y_i)
6. Total number of persons enumerated on long forms within each AA (Y_{Li})
7. AA Sampling Rate for each AA ($AASR_i$)
8. S^2 for each AA (S_i^2)
9. t
10. $t_{critical}$
11. The results (pass or fail) for the batch
12. The variance for the batch ($V(\hat{Y})$)

Resampling Output (for failing batches or where $SR_i < AASR_i$)

1. Take-every (TE_i)
2. Random start (r)
3. Generated sequence (L_x)
4. Random start (q) - *if necessary*
5. Generated sequence (L_z) - *if necessary*
6. List of IDs with original corresponding form type
7. List of IDs with corresponding form type after resampling